DOCUMENT RESUME

ED 336 394 TM 017 033

AUTHOR McCallister, Corliss
TITLE Phi, Rho, P.M., Biserial and Point-Biserial "r": A Review of Linkages.
PUB DATE Jan 91
NOTE 30p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 24-26, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; *Correlation; Educational Research; Error of Measurement; Psychological Studies; *Research Methodology; Robustness (Statistics); Scores
IDENTIFIERS Biserial Correlation; Heterogeneity of Variance; *Linkage Analysis; *Pearson Product Moment Correlation; Phi Coefficient; Point Biserial Correlation; Power (Statistics); Range Restriction; Rho Coefficient

ABSTRACT

The Pearson product moment (P.M.) correlation ("r") and four of its most widely used variations--the phi, the rho, the biserial, and the point-biserial coefficients--are reviewed. Using small data sets between one and nine, the conditions under which the various forms are restricted in power and robustness are explored. Seven sample data sets were constructed to illustrate the effects of the following conditions: (1) perfect correlation; (2) restriction of range; (3) measurement error (one variable); (4) measurement error (two variables); (5) extreme scores (one outlier); (6) extreme scores (two outliers); and (7) heterogeneity of sample distribution (one variable). Conceptual and algebraic linkages among the coefficients, and teaching practices that will facilitate the assimilation of these concepts and that will allow a student to predict the practical effects of the conditions are discussed. Two data tables are included. Appendix A contains definitions, formulas, and examples of the coefficients, which are presented as reproductions of a Macintosh computer hypercard stack designed for instructional purposes. Appendix B contains sample spreadsheet calculations for the seven conditions studied. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CORLISS MCCALLISTER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED336394

Phi, rho, p.m., biserial and point-biserial r: A review of linkages

Corliss McCallister

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX, January 26, 1991.

TM 017033

BEST COPY AVAILABLE

Abstract

Researchers in education and psychology can choose from an assortment of bivariate correlation coefficients depending upon the characteristics of the variables they measure and the instruments being used to measure them. These options and the conditions under which they can appropriately be employed are generally presented in statistics texts (Glass & Hopkins, 1984; Hinkle, Wiersma & Jurs, 1988). While most researchers and students are aware of the basic properties of r, many have not explored the conceptual and algebraic linkages among its various forms. Nor have they explored the influences commonly encountered in empirical studies on these alternative versions of the "simple" correlation coefficient (Busby & Thompson, 1990). This paper briefly reviews the many forms of r and explores, through the use of small data sets, the conditions under which various forms are restricted in power and robustness. For students to be adequately prepared for their roles as future researchers, statistics courses must not only provide effective instruction regarding the family of bivariate correlation coefficients and the conditions impacting its most advantageous use, but also the high costs of the unwise selection of coefficients (Cohen, 1983; Fowler, 1987). Suggestions for teaching these topics are also offered.

Phi, rho, p.m., biserial and point-biserial r:
A review of linkages

Correlation is defined as the index that describes the extent to which two sets of data are related (Hinkle, Wiersma & Jurs, 1988). The Pearson product-moment correlation coefficient (Pearson's r) is the best known of these indices; however, a number of variations of the Pearson coefficient have been developed to facilitate the analysis of data measured on nominal or ordinal scales. Although the family of bivariate correlation coefficients has many useful members, this paper will selectively focus on the Pearson r and four of its most widely used variations--the phi, the rho, the biserial and the point-biserial coefficients.

Definitions, formulas and examples of these coefficients are found in Appendix A. These pages are printed reproductions of a Macintosh microcomputer hypercard stack designed for instructional purposes. They represent the traditional approach to teaching these coefficients, one which emphasizes the differences between the appearances of the formulas rather than their commonalities. Conceptual linkages, though theoretically abundant, are not always explicitly presented by texts or lectures. While some textbooks (notably Hinkle, Wiersma & Jurs, 1988) do provide algebraic proofs for the derivation of these formulas, showing in fact that the formula are algebraically linked, even this exercise may not result in conceptual understanding for all students.

Application problems specifically requiring the calculation, inspection and comparison of the obtained coefficients using various

formulas may facilitate the realizations necessary to transferring knowledge from abstract presentations to the practical use of these coefficients in empirical settings. Homework problems, whether calculated by hand or by statistical packages such as SAS or SPSS-X, do little to help the student synthesize the five coefficients, if only one of the coefficients is computed for each data set. If the student lacks the final comprehension step of integrating these relationships, incomplete understanding may result in future misuse of one or more of the coefficients and, more regrettably, in faulty experimental designs producing sample statistics which dramatically underestimate the population parameter. Teaching practices which will facilitate the assimilation of these concepts and which will allow the student to predict the practical effects of the various conditions impacting on correlation coefficients will be discussed later in this paper.

Not only are these five coefficients linked conceptually and algebraically, they are also linked in their common requirements of the basic assumptions. Satisfying the assumptions of linearity and homoscedasticity (i.e., both variables have a common shape) is required for the robustness of the product moment correlation coefficient. Violations of these assumptions can lead to the reduction of the maximum r obtained. Few studies have explored the violation of assumptions for coefficients other than Pearson's r. Kraemer (1980) applied these results to the study of the point biserial and phi coefficients and found that these coefficients are likewise affected.

In the present paper seven data sets were constructed to illustrate the effects of various conditions on the five bivariate correlation coefficients being studied. These are presented in Table 1. One data set representing a perfect correlation and six data sets with variations representing conditions commonly described in the measurement literature were included and all five correlation coefficients were calculated for each set.

Insert Table 1 about here

The purpose of these sample data sets was to examine the effects of these various influences as they differentially affected specific coefficients. Listed below are the various conditions created:

(a) perfect correlation;

(b) restriction of range;

(c) measurement error (one variable);

(d) measurement error (both variables);

(e) extreme scores (one outlier);

(f) extreme scores (two outliers); and

(g) heterogeneity of sample distribution (one variable).

Appendix B presents the printout of spreadsheet software used to calculate the various coefficients for these seven datasets.
Table 2 presents the values of the five correlation coefficients for each of the conditions explored. The computed values for each of the six subsequent conditions drop below the theoretical value and the

computed value of Pearson's r for the perfect correlation in Dataset 1 and this phenomenon is especially pronounced for the phi coefficient in all cases. The investigation revealed that correlation coefficients were reduced in each of these six conditions and that the reductions differed by both condition and coefficient.

Insert Table 2 about here.

Whether interval or ratio data is subsequently rank ordered or dichotomized, a loss of information occurs. When the original data are inherent ranks (e.g., class ranks) or inherent dichotomies (e.g., gender), the researcher has no choice but to utilize the formula appropriate to such data. However in the event that normal distributions underlie the continuous variables being measured, and interval or ratio data might be collected, creating dichotomies is ill-advised (Cohen, 1982; Fowler, 1987). Busby explains that rank ordered data also will yield a smaller correlation than continuous data. According to Wylie (1976), "there is little question that correlations computed for such dichotomous variables provide very poor estimates of the true correlations between the continuous forms of these variables" (p. 5).

A major topic in measurement research which has both theoretical and applied interest is the attenuation of r values by a variety of conditions both under and outside of the control of the experimenter. Dataset 2 represents such a sample in which the range of both the x and y variables are severely restricted. This restriction of range will cause

attenuation of r in both direct and indirect ways according to Busby and Thompson (1990). Although the restriction of range resulting from the violations of the assumptions of linearity and homoscedasticity have received a great deal of attention in the measurement literature, other instances of this phenomenon arising as the effects of experimental design and research instrumentation have also been explored.

The problem of correcting correlation coefficients has also received a great deal of attention. Methods for approximating corrections in such situations have been offered for conditions of one variable restricted in some way, or for both variables when restricted. Alexander, Carson, Alliger and Carr (1987) presented a method for correcting doubly truncated correlations. Once a student has an impression of the influences restricting r, learning correction procedures will also be understood in terms of conceptual linkages.

Just as the validity of the tests establishes a ceiling for its reliability, the reliability of the tests establishes a ceiling for the correlation obtained between scores on that test and other experimental data. This reduction in r occurs because of the effect of measurement error on the standard deviation of a set of scores. Because measurement error is <u>random</u> error by definition, and because the standard deviation is used in the denominator, a test with low validity or reliability will result in data having a greater standard deviation consisting of proportionately more random error. This larger number in the denominator will result in a smaller answer. The reliability coefficients of the variables being correlated determine the extent of

measurement error, thus contributing to attenuation of the r. Despite the recognition of this problem, many students underestimate the extent of this reduction. To demonstrate the effects of measurement error on each of these coefficients, two datasets were constructed. Datasets 3 and 4 illustrate the effect of measurement error on five correlations for one variable and two variables being affected, respectively.

Halperin (1986) showed that when variables have marginal distributions that have heavy tails, outliers are produced. As a result of these extreme scores, correlations may be either spuriously large or small. According to Halperin, exploratory data analysis should be used to diagnose the problem of extreme scores and should be followed by corrective actions (such as selectively discarding discordant observations). An example of the effect of extreme scores is shown in Dataset 5. One outlier was substituted for an original value; reductions in all the coefficients resulted. A further illustration of this effect is given in Dataset 6 in which two outliers have been used in place of original data points.

According to Wylie (1976), it has been recognized among researchers for a considerable period of time that "the product moment correlation coefficient is severely restricted in the values it can assume when marginal distributions are both dichotomous and asymmetrical" (p. 1). While the robustness of the Pearson r has been examined both analytically and empirically, less extensive investigations have been performed on the robustness of the other coefficients. In a Monte Carlo

study performed by Fowler (1987), the power of the point-biserial and phi coefficients were determined using relatively small samples drawn from bivariate populations of various shapes. Fowler reported that although the point biserial r showed "good relative power (>120%) for some distributional shapes" compared to Pearson's r, substantial reduction in power was demonstrated for other conditions especially with small sample sizes. The phi coefficient, also studied under these varying distribution shapes showed poor relative power "for all population shapes under eleven conditions simulated" (p. 425). Therefore, when individual distributions are of differing shapes, the maximum possible correlations are reduced, as seen in Dataset 7.

As correlation coefficients may well be the single most utilized statistical measure in educational and psychological research, any additional time and energy devoted to these decisions could easily be justified. Yet the suggestions for activities which would hothouse such concept development and practical knowledge are neither excessively time-consuming nor energy-consuming.

A simple example of one of these teaching methods which results in both the conceptual linkage and the practical experience desired is the spread sheet. One advantage of using the spreadsheet is its speed and ease of calculation, once the software has been prepared. Two further advantages to using the spreadsheet application to teach correlation coefficients are the direct control over the speed and complexity of the process by the student and the immediate visual display of changes in each of the statistics computed. How better could

students learn that the correlations among such variables are inextricably linked? How much more easily could they monitor the effect on r by manipulations of the data set? What better way is there to learn that other variables of interest may be substantially restricted?

Suggested activities for the teaching of the bivariate family of correlation coefficients, which promote understanding of the linkage of the various forms of Pearson's r, are listed below. In each activity, each student or group of students in the class can work independently on datasets of their choosing and then share the correlation coefficients with classmates to spark generalizations concerning the trends apparent in their data.

Activity 1. Creation of data sets to illustrate the conditions impacting on correlation coefficients with a given effect size similar to Table 1.

Activity 2. Completion of a table of correlations similar to Table 2 which would emphasize the magnitude of the restrictions caused by various conditions (e.g., dichotomization).

Activity 3. Comparison of correlation coefficients for random sampling subsets of a known population.

Activity 4. Preparation of alternative hypercard lessons similar in format to Appendix A but emphasizing linkages among the correlation coefficients.

Activity 5. Preparation of spreadsheets similar to Appendix B for the calculation of one or all of the correlation coefficients.

As bivariate correlation coefficients are commonly used in both educational and psychological research, students being trained for these professions need effective instruction both in the various forms that these coefficients can take and the conditions which affect these coefficients. Such instruction can result in the most advantageous choices of data gathering and data analysis procedures, and in the correct statistical interpretation of the experimental results.

References

Alexander, R. L., Carson, K. P., Alliger, G. M. & Carr, L. (1987). Correlating doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. Educational and Psychological Measurement, 47(2), 309-315.

Busby, D. & Thompson, B. (1990, January). Factors attenuating Pearson's r: A review of basics and some corrections. Paper presented at the annual meeting of the Southwestern Educational Research Association, Austin, TX.

Cohen, J. (1983). The cost of dichotomization. Applied Psychological Measurement, 7(3), 219-253.

Fowler, R. L. (1987). Power and robustness in product-moment correlation. Applied Psychological Measurement, 11(4), 419-428.

Glass, G. V. & Hopkins, K. D. (1984). Statistical methods in education and psychology. Englewood Cliffs: Prentice-Hall.

Halperin, S. (1986). Spurious correlations: Causes and cures. Psychoneuroendocrinology, 11(1), 3-13.

Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1988). Applied statistics of the behavioral sciences. Boston: Houghton-Mifflin, 521-544.

Kraemer, H. C. (1980). Robustness of the distribution theory of the product moment correlation coefficient. Journal of Educational Statistics, 5(2), 115-128.

Wylie, P. B. (1976). Effects of coarse grouping and skewed marginal distributions on the Pearson product moment correlation coefficient. Educational and Psychological Measurement, 36, 1-7.

Table 1

Seven Datasets Constructed to Illustrate Conditions Impacting on the Five Types of Correlation Coefficients Studied

Dataset 1. Perfect Correlation

1,1 2,2 3,3 4,4 5,5 6,6 7,7 8,8 9,9

Dataset 2. Restricted Range

1,1 2,2 3,3 1,1 2,2 3,3 1,1 2,2 3,3

Dataset 3. Measurement Error(One Variable)

1.60, 1 1.17, 2 3.32, 3 3.81, 4 5.11, 5 5.69, 6 7.88, 7 7.70, 8 9.22, 9

Dataset 4. Measurement Error (Two Variables)

1.16, 0.18 1.17, 2.82 3.32, 2.93 3.81, 4.05 5.11, 4.17 5.69, 6.98 7.88, 6.99 7.70, 8.63 9.22, 9.07

Dataset 5. Extreme Scores (One Outlier)

1,1 2,2 3,3 4,4 5,5 6,6 7,7 8,3 9,9

Dataset 6. Extreme Scores (Two Outliers)

1,9 2,2 3,3 4,4 5,5 6,6 7,7 8,3 9,9

Dataset 7. Heterogeneous Distributions (One Variable Skewed)

3,4 4,1 4,2 5,1 5,2 5,3 6,1 6,3 7,5

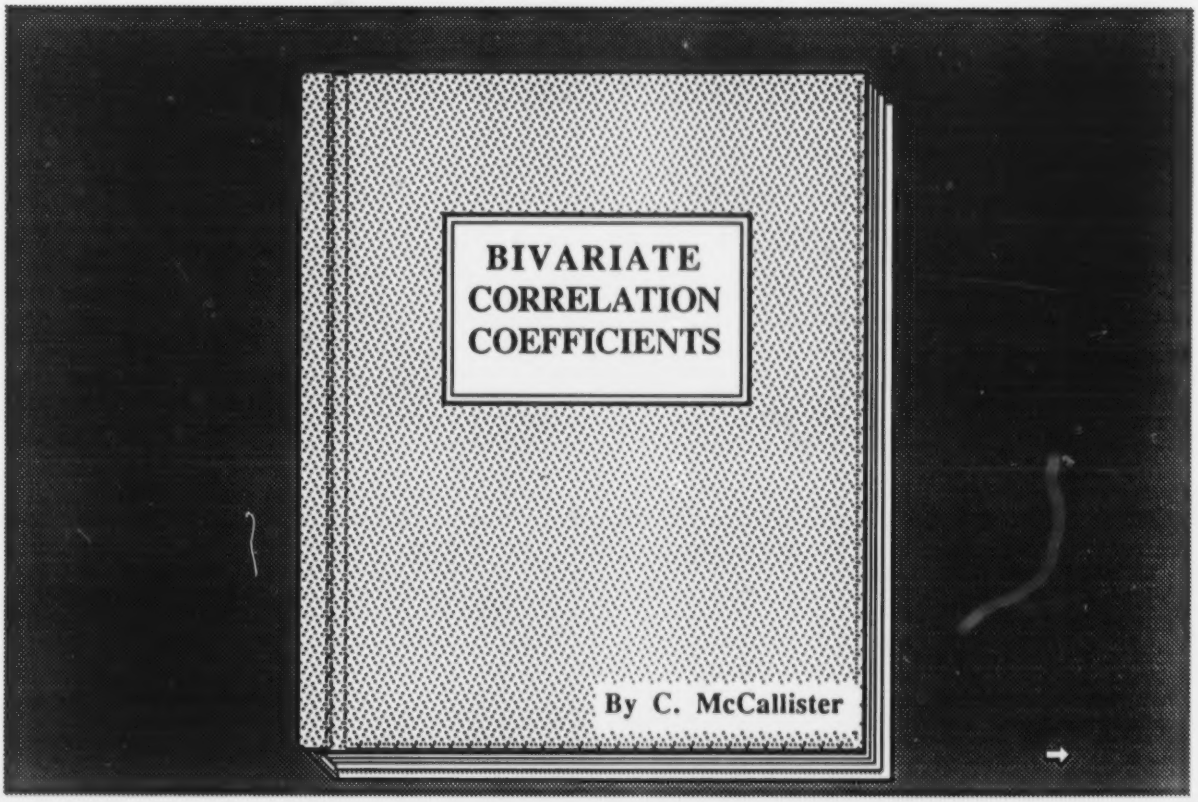
Table 2

Summary of Obtained Correlations for Seven Conditions and Five Coefficients of Correlation

	r	ϕ	ρ	r_b	r_{pb}
Dataset 1	1.00	1.00	1.00	0.87	0.87
Dataset 2	1.00	0.00	1.00	0.00	0.00
Dataset 3	0.98	1.00	0.97	0.86	0.87
Dataset 4	0.93	0.80	0.97	0.86	0.82
Dataset 5	0.80	0.80	0.79	0.64	0.72
Dataset 6	0.23	0.55	0.26	0.26	0.30
Dataset 7	0.22	0.25	0.16	0.61	0.06

Appendix A

Instructional Hypercard Stack



Introduction: Bivariate Correlation Coefficient

An index that describes the extent to which two sets of data are related;
it is a measure of the relationship of two variables.

The appropriate correlation coeffient depends on the scales of measurement of the two variables being correlated.

EXAMPLES:

Phi	ϕ
Rho	ρ
Product Moment	r
Biserial	r_b
Point-Biserial	r_{pb}

PEARSON'S PRODUCT MOMENT
(PEARSON'S r)

AN INDEX OF THE LINEAR RELATIONSHIP BETWEEN TWO VARIABLES.

USED APPROPRIATELY WHEN:
(1) BOTH VARIABLES ARE MEASURED ON EITHER AN INTERVAL OR A RATIO SCALE AND
(2) THE UNDERLYING DISTRIBUTIONS OF BOTH VARIABLES ARE NORMAL.

EXAMPLES: HEIGHT,WEIGHT

FORMULA:

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

RANGE:
-1.00 to 1.00

PHI (ϕ)

A special case of the Pearson r in which both variables are nominal dichotomies.

Examples: men or women, Republican or Democrat.

Formula:

$$\phi = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

Range: -1.00 to 1.00

RHO (ρ)

A special case of the Pearson r that is used when the scores on the two variables being correlated are ranks.

Examples: Rank in graduating class, ranked SAT-V scores.

Formula:

$$\rho = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

Range:

-1.00 to 1.00

POINT BISERIAL

A special case of the Pearson r, in which one variable is a quantitative variable measured on an interval or ratio scale and the other is a nominal (dichotomous) variable.

EXAMPLE: IQ, GENDER

FORMULA:

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{\sigma_Y} \sqrt{pq}$$

RANGE:

-1.00 to 1.00

BISERIAL (r_b)

A special case of the Pearson r which is appropriate when one variable is quantitative and measured on the interval or ratio scale and the other variable is dichotomous but with underlying continuity.

EXAMPLE: creativity, gifted or not

FORMULA:

$$r_b = \left(\frac{\bar{Y}_1 - \bar{Y}_0}{\sigma_Y} \right) \left(\frac{pq}{y} \right)$$

RANGE:

Theoretically -1.00 to 1.00
However, rarely, can exceed these values.

Appendix B

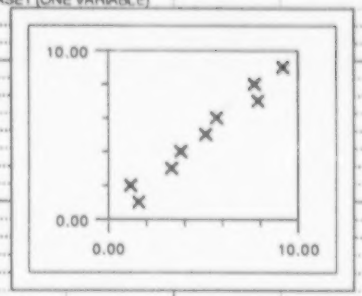
Sample Spreadsheet Calculations

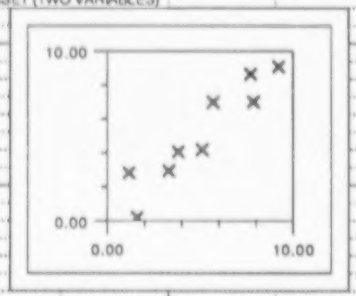
BIVARIATE CORRELATION COEFFICIENTS

PERFECT CORRELATION DATASET

| | | |

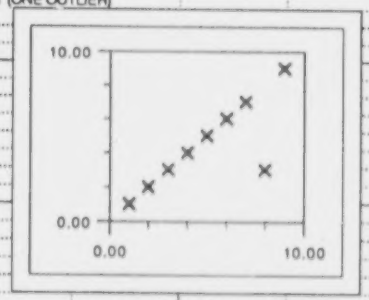
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
55	MEASUREMENT ERROR DATASET (ONE VARIABLE)						X^2	Y^2	XY	v1	v2	RANK X	RANK Y	D	D^2	MEAN Y 1	MEAN Y 0	STD DEV Y	P * Q	ORDINATE HT	A	B	C	D
56	1.60	1.00					2.56	1.00	1.60	0	0	2.00	1.00	1.00	1.00	5.00	1.00	1.00	0.2469	0.3953	0	0	1	0
57	1.17	2.00					1.37	4.00	2.34	0	0	1.00	2.00	-1.00	1.00	8.00	2.00	2.00			0	0	1	0
58	3.32	3.00					11.02	9.00	9.96	0	0	3.00	3.00	0.00	0.00	7.00	3.00	3.00			0	0	1	0
59	3.81	4.00					14.52	16.00	15.24	0	0	4.00	4.00	0.00	0.00	6.00	4.00	4.00			0	0	1	0
60	5.11	5.00					26.11	25.00	25.55	1	0	5.00	5.00	0.00	0.00	9.00		5.00			0	0	0	1
61	5.69	6.00					32.38	36.00	34.14	1	1	6.00	6.00	0.00	0.00			6.00			0	1	0	0
62	7.88	7.00					62.09	49.00	55.16	1	1	8.00	7.00	1.00	1.00			7.00			0	1	0	0
63	7.70	8.00					59.29	64.00	61.60	1	1	7.00	8.00	-1.00	1.00			8.00			0	1	0	0
64	9.22	9.00					85.01	81.00	82.98	1	1	9.00	9.00	0.00	0.00			9.00			0	1	0	0
65	45.50	45.00					294.35	265.00	286.57						4.00	5.00	4.00	45.00			0	4	4	1
66																7.00	2.50	2.58199						
67	9	9																						
68																								
69																								
70																								
71																								
72																								
73	P.M. =	0.98																						
74	PHI =	0.80																						
75	RHO =	0.97																						
76	BISERIAL =	1.09																						
77	POINT BIS =	0.87																						
78																								
79																								
80	MEASUREMENT ERROR DATASET (TWO VARIABLES)						X^2	Y^2	XY	v1	v2	RANK X	RANK Y	D	D^2	MEAN Y 1	MEAN Y 0	STD DEV Y	P * Q	ORDINATE HT	A	B	C	D
81	1.60	0.18					2.56	0.03	0.29	0	0	2.00	1.00	1.00	1.00	4.17	0.18	0.18	0.2469	0.3953	0	0	1	0
82	1.17	2.82					1.37	7.95	3.30	0	0	1.00	2.00	-1.00	1.00	6.87	2.82	2.82			0	0	1	0
83	3.32	2.93					11.02	8.58	9.73	0	0	3.00	3.00	0.00	0.00	6.99	2.93	2.93			0	0	1	0
84	3.81	4.05					14.52	16.40	15.43	0	0	4.00	4.00	0.00	0.00	8.83	4.05	4.05			0	0	1	0
85	5.11	4.17					26.11	17.39	21.31	1	0	5.00	5.00	0.00	0.00	9.07		4.17			0	0	0	1
86	5.69	6.98					32.38	48.72	39.72	1	1	6.00	6.00	0.00	0.00			6.87			0	1	0	0
87	7.88	8.99					62.09	48.86	55.08	1	1	8.00	7.00	1.00	1.00			8.99			0	1	0	0
88	7.70	8.63					59.29	74.48	66.45	1	1	7.00	8.00	-1.00	1.00			8.63			0	1	0	0
89	9.22	9.07					85.01	82.26	83.63	1	1	9.00	9.00	0.00	0.00			9.07			0	1	0	0
90	45.50	45.82					294.35	304.68	294.93						4.00	5.00	4.00	45.71			0	4	4	1
91																7.15	2.50	2.81679						
92	9	9																						
93																								
94																								
95																								
96																								
97																								
98	P.M. =	0.93																						
99	PHI =	0.80																						
100	RHO =	0.97																						
101	BISERIAL =	1.03																						
102	POINT BIS =	0.82																						
103																								
104																								

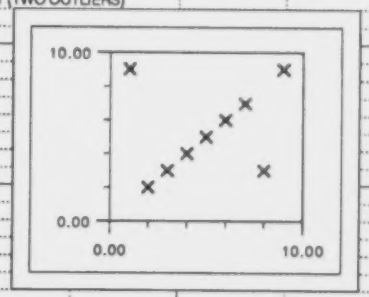




BEST COPY AVAILABLE

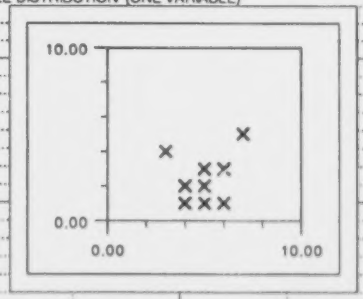
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
105	EXTREME SCORES DATASET (ONE OUTLIER)						X^2	Y^2	XY	v1	v2	RANK X	RANK Y	D	D^2	MEAN Y 1	MEAN Y 0	STD DEV Y	P * Q	ORDINATE HT	A	B	C	D
106	1.00	1.00					1.00	1.00	1.00	0	0	1.00	1.00	0.00	0.00	6.00	1.00	1.00	0.2469	0.3953	0	0	1	0
107	2.00	2.00					4.00	4.00	4.00	0	0	2.00	2.00	0.00	0.00	7.00	2.00	2.00			0	0	1	0
108	3.00	3.00					9.00	9.00	9.00	0	0	3.00	3.50	-0.50	0.25	3.00	3.00	3.00			0	0	1	0
109	4.00	4.00					16.00	16.00	16.00	0	0	4.00	5.00	-1.00	1.00	9.00	4.00	4.00			0	0	1	0
110	5.00	5.00					25.00	25.00	25.00	0	0	5.00	6.00	-1.00	1.00		5.00	5.00			0	0	1	0
111	6.00	6.00					36.00	36.00	36.00	1	1	6.00	7.00	-1.00	1.00			6.00			0	1	0	0
112	7.00	7.00					49.00	49.00	49.00	1	1	7.00	8.00	-1.00	1.00			7.00			0	1	0	0
113	8.00	3.00					64.00	9.00	24.00	1	0	8.00	3.50	4.50	20.25			3.00			0	0	0	1
114	9.00	9.00					81.00	81.00	81.00	1	1	9.00	9.00	0.00	0.00			9.00			0	1	0	0
115	45.00	40.00					285.00	230.00	245.00						24.50	4.00	5.00	40.00			0	3	5	1
116																6.25	3.00	2.40883						
117	9	9																						
118																								
119																								
120																								
121																								
122																								
123	P.M. =	0.80																						
124	PHI =	0.79																						
125	PHO =	0.80																						
126	BISERIAL =	0.84																						
127	POINT BIS =	0.67																						
128																								
129																								
130	EXTREME SCORES DATASET (TWO OUTLIERS)						X^2	Y^2	XY	v1	v2	RANK X	RANK Y	D	D^2	MEAN Y 1	MEAN Y 0	STD DEV Y	P * Q	ORDINATE HT	A	B	C	D
131	1.00	9.00					1.00	81.00	9.00	0	1	1.00	8.50	-7.50	56.25	6.00	9.00	9.00	0.2469	0.3953	1	0	0	0
132	2.00	2.00					4.00	4.00	4.00	0	0	2.00	1.00	1.00	1.00	7.00	2.00	2.00			0	0	1	0
133	3.00	3.00					9.00	9.00	9.00	0	0	3.00	2.50	0.50	0.25	3.00	3.00	3.00			0	0	1	0
134	4.00	4.00					16.00	16.00	16.00	0	0	4.00	4.00	0.00	0.00	9.00	4.00	4.00			0	0	1	0
135	5.00	5.00					25.00	25.00	25.00	0	0	5.00	5.00	0.00	0.00		5.00	5.00			0	0	1	0
136	6.00	6.00					36.00	36.00	36.00	1	1	6.00	6.00	0.00	0.00			6.00			0	1	0	0
137	7.00	7.00					49.00	49.00	49.00	1	1	7.00	7.00	0.00	0.00			7.00			0	1	0	0
138	8.00	3.00					64.00	9.00	24.00	1	0	8.00	2.50	5.50	30.25			3.00			0	0	0	1
139	9.00	9.00					81.00	81.00	81.00	1	1	9.00	8.50	0.50	0.25			9.00			0	1	0	0
140	45.00	48.00					285.00	310.00	253.00						56.00	4.00	5.00	48.00			1	3	4	1
141																6.25	4.60	2.44949						
142	9	9																						
143																								
144																								
145																								
146																								
147																								
148	P.M. =	0.23																						
149	PHI =	0.55																						
150	PHO =	0.27																						
151	BISERIAL =	0.42																						
152	POINT BIS =	0.33																						
153																								
154																								





BEST COPY AVAILABLE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
155	HETEROGENEITY OF SAMPLE DISTRIBUTION (ONE VARIABLE)						X^2	Y^2	XY	v1	v2	RANK X	RANK Y	D	D^2	MEAN Y 1	MEAN Y 0	STD DEV Y	P * Q	ORDINATE HT	A	B	C	D
156	3.00	4.00					9.0000	16.0000	******	0	0	1.00	8.00	-7.00	49.00	1.00	4.00	4.00	0.2222	0.3643	0	0	1	0
157	4.00	1.00					16.0000	1.0000	4.0000	0	0	2.50	2.00	0.50	0.25	3.00	1.00	1.00			0	0	1	0
158	4.00	2.00					16.0000	4.0000	8.0000	0	0	2.50	4.50	-2.00	4.00	5.00	2.00	2.00			0	0	1	0
159	5.00	1.00					25.0000	1.0000	5.0000	0	0	5.00	2.00	3.00	9.00		1.00	1.00			0	0	1	0
160	5.00	2.00					25.0000	4.0000	******	0	0	5.00	4.50	0.50	0.25		2.00	2.00			0	0	1	0
161	5.00	3.00					25.0000	9.0000	******	0	0	5.00	6.50	-1.50	2.25		3.00	3.00			0	0	1	0
162	6.00	1.00					36.0000	1.0000	6.0000	1	0	7.50	2.00	5.50	30.25			1.00			0	0	0	1
163	6.00	3.00					36.0000	9.0000	******	1	0	7.50	6.50	1.00	1.00			3.00			0	0	0	1
164	7.00	5.00					49.0000	25.0000	******	1	0	9.00	9.00	0.00	0.00			5.00			0	0	0	1
165	45.00	22.00					*******	70.0000	******						96.00	3.00	6.00	22.00			0	0	6	3
166																3.00	2.17	1.34256						
167	9	9																						
168																								
169																								
170																								
171																								
172																								
173	P.M. =	0.22																						
174	PHI =	ERR 9																						
175	PHO =	0.20																						
176	BISERIAL =	0.38																						
177	POINT BIS =	0.29																						
178																								
179																								



BEST COPY AVAILABLE